Conservation Biology

Contributed Paper

Estimating How Many Undescribed Species Have Gone Extinct

P. A. TEDESCO,* R. BIGORNE,* A. E. BOGAN,† X. GIAM,‡§ C. JÉZÉQUEL,* AND B. HUGUENY*

*UMR Biologie des Organismes et des Ecosystèmes Aquatiques (MNHN-IRD-CNRS-UPMC), Muséum National d'Histoire Naturelle, 43 rue Cuvier–CP 26, 75005, Paris, France, email pablo.tedesco@mnhn.fr
†North Carolina State Museum of Natural Sciences, 11 West Jones St, Raleigh, NC, 27601, U.S.A.
‡Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, 08544, U.S.A.
§Raffles Museum of Biodiversity Research, National University of Singapore, S117543, Singapore

Abstract: *Because both descriptions of species and modern human-driven extinctions started around the same time (i.e., eighteenth century), a logical expectation is that a large proportion of species may have gone extinct without ever having been recorded. Despite this evident and widely recognized assumption, the loss of undescribed species has never been estimated. We quantified this loss for several taxonomic groups and regions for which undescribed species extinctions are likely to have occurred. Across a wide range of taxonomic groups, we applied known extinction rates computed from recorded species losses to assumed exponential decay in the proportion of species remaining undiscovered. Because all previous modeling attempts to project total species richness implicitly assumed that undescribed species extinctions could be neglected, we also evaluated the effect of neglecting them. Finally, because we assumed constant description and extinction probabilities, we applied our model to simulated data that did not conform to this assumption. Actual species losses were severely underestimated by considering only known species extinctions. According to our estimates, the proportion of undiscovered extinct species over all extinctions ranged from 0.15 to 0.59, depending on the taxonomic group and the region considered. This means that recent extinctions may be up to twice as large as the number recorded. When species differed in their extinction or description probabilities, our model underestimated extinctions of undescribed species by up to 20%.*

Keywords: conservation, description curve, diversity, extinction, taxonomy

Estimación de Cuántas Especies sin Describir Se Han Extinguido

Resumen: *Ya que la descripción de especies y las extinciones causadas por el humano iniciaron aproximadamente al mismo tiempo (es decir, en el siglo XVIII), una expectativa lógica es que una gran proporción de las especies puede haberse extinguido sin siquiera haber sido registrada. A pesar de esta suposición evidente y ampliamente reconocida, nunca se ha estimado la pérdida de especies sin describir. Cuantificamos esta pérdida para varios grupos y regiones taxonómicas para las cuales es probable que hayan ocurrido extinciones de especies sin describir. A lo largo de un amplio rango de grupos taxonómicos, aplicamos tasas de extinción conocidas obtenidas de cómputos de pérdidas registradas de especies al declive exponencial en la proporción de especies que permanecen sin ser descubiertas. Como todos los intentos previos de modelos para proyectar la riqueza total de especies suponían implícitamente que la extinción de especies sin describir podía negarse, también evaluamos el efecto de negarlos. Finalmente, ya que supusimos las probabilidades de extinción y de descripción constante, aplicamos nuestro modelo a datos simulados que no se conforman a esta suposición. Las pérdidas recientes de especies estuvieron severamente subestimadas considerando solamente las extinciones de especies conocidas. De acuerdo con nuestras estimaciones, la proporción de especies extintas no descubiertas en todas las extinciones abarcó desde 0.15 a 0.59, dependiendo del grupo taxonómico y la región considerada. Esto significa que las extinciones recientes pueden ser hasta dos veces mayores al número registrado. Cuando las especies difirieron en su extinción o en la probabilidad de descripción, nuestro modelo subestimó la extinción de especies sin describir por hasta 20%.*

Palabras Clave: conservación, curva de descripción, diversidad, extinción, taxonomía

Paper submitted September 24, 2012; revised manuscript accepted January 11, 2014.

© 2014 Society for Conservation Biology
DOI: 10.1111/cobi.12285

Introduction

Rates of species extinction due to human actions are considered higher than geologically typical rates (Pereira et al. 2010) and could even reach levels comparable to past mass extinctions in the next centuries if current threats to species are not alleviated (Barnosky et al. 2011). Moreover, documented numbers of current species extinctions are likely to be serious underestimates because most species have not yet been formally described (e.g., Stork 1993; Mora et al. 2011; Scheffers et al. 2012), even for charismatic groups otherwise considered well known to scientists (Ceballos & Ehrlich 2009). Because both descriptions and modern human-driven extinctions started at nearly the same time (i.e., eighteenth century), a logical prediction is that an unknown proportion of species may have gone extinct without ever being recorded. Furthermore, because new species continue to be described and human-driven threats are not expected to level off, in the future more species are likely to go extinct without being noticed. Hence, the already severe current extinction crisis (Pereira et al. 2010; Stork 2010; Barnosky et al. 2011) and the ongoing rise of species endangerment (IUCN 2012) are likely worse than documented. Although the extinction of a species unknown to science is, in essence, not verifiable, recent descriptions based on remains of species that were never collected alive but were likely extinct following recent human settlement are tangible proof that extinctions of undescribed species occurred (e.g., Richling & Bouchet 2013).

Despite this evidence (e.g., Ceballos & Ehrlich 2009; Giam et al. 2012; Scheffers et al. 2012), few estimations of the undescribed diversity loss have been attempted. Most attempts focused on Pacific island birds that went extinct after their first contact with humans (before the sixteenth century) without leaving remains (e.g., bones) that would have allowed subsequent formal description (e.g., Pimm et al. 1994). We focused on unnoticed extinctions that occurred because naturalists began to describe species with binomial nomenclature. We attempted to quantify this loss for several taxonomic groups for which and regions where undescribed species extinctions are likely to have occurred. We modeled species description trends (Solow & Smith 2005; Giam et al. 2010; Joppa et al. 2011*a*) and applied known species extinction rates to the predicted number of undescribed species.

We assumed that the modern drivers leading species to extinction (e.g., habitat destruction, fragmentation, overexploitation, nonnative species introduction) have affected similarly both described and undescribed species through time. Although this assumption seems reasonable, it should be considered conservative because as the description process proceeds, species waiting to be discovered are more likely to be range restricted (Scheffers et al. 2012), a species trait widely associated with increased likelihood of extinction (e.g., McKinney 1997; Purvis et al. 2000*a*, 2000*b*). For instance, the positive relationship between extinction proneness of species and year of description, which suggests that many undescribed species are in danger of extinction, is mainly driven by restricted range species (although not exclusively) (Giam et al. 2012). Our estimates of undescribed extinct species are thus likely to be underestimates. Because we assumed that species do not differ with regard to their probability of description and extinction, we explored the potential impact of these simplifying assumptions on our results via simulations.

All models of species discovery (e.g., Solow & Smith 2005; Bebber et al. 2007; Joppa et al. 2011*a*) are based implicitly on the assumption that extinctions of undescribed species do not affect a model's projections of total species richness. We evaluated this assumption by comparing the outcomes of our species description model when accounting or not accounting for undescribed species extinctions.

Methods

The model

We modeled species discovery curves assuming that described and undescribed species went extinct with an annual probability e and that undescribed species have been described with annual probability d. The description probability per species per year (d) depends on several interacting factors related to the taxonomic research process (e.g., effort expended in discovering new species, expertise in identifying new species, and proportion of habitat remaining unexplored) and to species intrinsic characteristics (e.g., their visibility). Because we had no means of separating and estimating these factors, we considered the simplest model of the discovery record in which variation in these factors is low compared with the effect of diminishing new species and d is therefore constant (Bebber et al. 2007). We assumed that natural extinction probability is so low that it could be neglected. Therefore, the extinction probability e refers to recent, human-driven extinctions.

For clarity, we first describe a model where undescribed species extinctions are not occurring (i.e., decaying number of undescribed species diminishes through time due to descriptions only). We then added undescribed species extinctions to the model to finally compute the cumulated number of undescribed species that have gone extinct.

We defined t_d as the year at which a species is described. For simplicity, time is scaled such that $t_d = 1$ for the most ancient species' descriptions. Assuming d is constant through time in the absence of extinction, the

probability of a species being described at a given year t is

$$\mathrm{P}(t_d = t) = (1-d)^{t-1}d. \quad (1)$$

We defined the random variable D_t as the number of species described during year t, and we postulated that description probability is constant among species. The expected value of D_t is then

$$\mathrm{E}(D_t) = N_0(1-d)^{t-1}d, \quad (2)$$

where N_0 is the initial number of species before starting the description process. According to geometric series and defining x as the common ratio of a sequence where $x = 1 - d$ and dN_0 is the scale factor, the cumulated number of species described over n years is

$$\sum_{t=1}^{t=n} \mathrm{E}(D_t) = dN_0\left[\frac{1-x^n}{1-x}\right]. \quad (3)$$

As expected, this cumulated number of described species tends to N_0 when n tends to infinity.

We considered that species extinctions occur simultaneously with descriptions with probability e. When descriptions and extinctions start at the same time ($t = 1$), to be available for description at year t a species cannot have been described or be extinct. Therefore, the probability of being described at year t is given by

$$P(t_d = t) = (1-e)^{t-1}(1-d)^{t-1}d. \quad (4)$$

A close approximation of Eq. 4 when e and d are $<<1$, which is the situation we examined, is

$$P(t_d = t) = (1-d-e)^{t-1}d. \quad (5)$$

Hence, we obtained a model of the discovery record that accounts for species extinction before being described:

$$E(D_t) = N_0(1-d-e)^{t-1}d \quad (6)$$

and

$$\sum_{t=1}^{t=n} \mathrm{E}(D_t) = dN_0\left[\frac{1-y^n}{1-y}\right] \quad (7)$$

where $y = 1 - (e + d)$.

We defined t_e as the year at which an undescribed species goes extinct. The probability of this event is

$$P(t_e = t) = (1-e)^{t-1}(1-d)^{t-1}e. \quad (8)$$

Using the same reasoning as in Eqs. 5–7, we estimated the cumulative number of undescribed species that have gone extinct with a geometric series:

$$\sum_{t=1}^{t=n} \mathrm{E}(E_t) = eN_0\left[\frac{1-y^n}{1-y}\right] = NE. \quad (9)$$

The total number of extinct species (both described and undescribed) over the period 0 to n is

$$TE = N_0[1-(1-e)^n]. \quad (10)$$

From Eqs. 9 and 10 we computed pE, the proportion of undescribed extinct species among all extinct ones over the period of interest (NE/TE) as

$$pE = \frac{e}{e+d}\frac{1-(1-d-e)^n}{1-(1-e)^n} \quad (11)$$

Then, if $e << d$ and $d << 1$, then an approximation of $(1-e)^n$ is $1 - ne$, and this relationship is reduced to

$$pE = \frac{1-(1-d)^n}{nd} \quad (12)$$

and leads to the counter-intuitive result that changes in extinction probability have no effect on the proportion of undescribed extinct species.

Suppose now that the first human driven species extinction occurs at year $t = m$, after the first species description. Because we assumed that natural extinction could be neglected, we considered that no species went extinct between years 0 and m and that CD_m species were described in that period. We define N_m as $N_0 - CD_m$. The N_0 should then be replaced by N_m in all equations excepting Eq. 10, and the right side of Eqs. 11 and 12 should be multiplied by N_m/N_0 to rescale pE values.

Parameter estimation

To run the model, we estimated 3 parameters: N_0, d, and e. We estimated e based on the assumption that extinction probability for described and undescribed species is equal. We therefore estimated extinction probability e from modern known species extinctions in each taxonomic group and region (for the list of considered species and sources, see Supporting Information). A very accurate approximation of e (if $e << 1$, see Supporting Information for details on this approximation) is given by

$$e = NE_{\mathrm{obs}} \Big/ \sum_{t=m}^{t=z} CD_{t,\mathrm{obs}}, \quad (13)$$

where NE_{obs} is the number of extinct species recorded over the period when these extinctions were observed, $CD_{t,\mathrm{obs}}$ is the cumulated number of described species to year t, and m and z are the years of first and last recorded extinctions, respectively. Use of the first and last recorded extinction dates (Table 1) to estimate e ensures that species considered extinct within that period are related to the number of species described within the same period and provides a per species extinction probability. Year m (i.e., year of the first recorded extinction) (Supporting Information) was used as a starting point to run our model. Although we have no means of knowing when extinctions of undescribed species began,

Table 1. Input information applied to run our model of species discovery curves for each taxonomic group and region considered.

Taxonomic group (reference code)[a]	*Region*	*Number of described species (year at $t = n$)*[b]	*Number of species considered extinct NE_{obs} (reference code)*[a]	*First (m) and last (z) years of observed extinctions*	*Extinction rate e ($sp^{-1}.yr^{-1}$)*	*Supporting Information code*
Freshwater fishes (a)	North America (U.S.A., Canada, Mexico)	1593 (2009)	41 (e, f, g)	1880–1994	0.0003071	1
Freshwater bivalves (b, c)	North America (U.S.A., Canada)	302 (2010)	28 (b, e, h, i)	1900–1990	0.0010940	2
Birds (d)	world (continental)	6183 (2010)	8 (e)	1852–1986	0.0000107	3
	Australia	628 (2010)	3 (e)	1805–1927	0.0000495	4
Amphibians (e)	Australia	223 (2007)	3 (e)	1979–1985	0.0024834	5
Marine mammals (e)	world	135 (2005)	4 (e, j)	1768–1955	0.0002490	6
Terrestrial mammals (e)	world (continental)	3950 (2010)	11 (e, j)	1894–1991	0.0000353	7
	Australia	292 (2008)	19 (e, k, j)	1840–1972	0.0007645	8

[a] *References for described and extinct species lists: a, Miller (2005), Tedesco et al. (2010); b, Williams et al., unpublished data; c, Turgeon et al. (1998); d, BirdLife International and NatureServe (2011); e, IUCN (2012); f, Harrison and Stiassny (1999); g, Jelks et al. (2008); h, Haag (2009); i, Régnier et al. (2009) ; j, MacPhee and Flemming (1999); k, Burbidge et al. (2009).*

[b] *Year of the last described species in our data set.*

the first known extinction date of a described species seems a reasonable guess. To analyze the influence of the starting year on our results, and because extinctions of undescribed species could have occurred earlier, we also ran our model starting 10 and 25 years before the first known extinction.

Although we estimated e over the period when extinctions were observed (i.e., from years m to z), we assumed extinctions continue at probability e after year z (i.e., until year n). We assumed that d and e are constant through time and species independent and that e equally affects both described and undescribed species. After setting parameter e from known extinctions, we estimated the remaining parameters d and N_0 by fitting Eq. 6 to D_t, the observed numbers of species descriptions per year. We used maximum likelihood, assuming that numbers of descriptions D_t follow a Poisson distribution (in agreement with our assumptions that d is very small and constant and that the number of described species and hence N_0 is large in our data sets), to estimate d and N_0. To minimize negative log likelihood, we applied the function *nlminb* from the R statistical package (R Development Core Team 2011). Finally, knowing all 3 parameters, we used Eq. 9 to estimate the number of species that have gone extinct without being described.

Species description and extinction data

We applied our model to records of discovery times from several taxonomic groups globally or regionally depending on the group (Table 1). Inclusion of a data set in our analysis depended on 3 conditions. First, the presence of a leveling off in the species description curve was necessary to reach accurate estimates of N_0 and d (i.e., to build the model from Eq. 6). Second, species extinctions must have been observed in the group within the considered region so that extinction probability could be estimated. Third, estimated extinction dates (e.g., last recorded occurrence of an extinct species) must be available from the literature or from databases in the public domain (e.g., International Union for Conservation of Nature; IUCN 2012; Committee on Recently Extinct Organisms [CREO]; Harrison & Stiassny 1999; MacPhee & Flemming 1999) because we needed a period within which extinction probability (e) could be estimated and a starting year to run our model. The taxonomic groups and regions meeting these criteria and the source information are in Table 1. When different information sources about extinct species were available (i.e., North American freshwater bivalves and fish and world terrestrial and marine mammals), we compiled all the information in one list of extinct species (Supporting Information). Using the different information sources separately did not change our overall results.

Recorded extinctions are not evenly distributed around the globe. For instance, a large proportion of mammal extinctions occurred in Australia (MacPhee & Flemming 1999; IUCN 2012). The same is true for freshwater bivalves in North America (Bogan 1993, 1998). For this reason, we distinguished between some geographic regions to provide more realistic extinction probabilities (Table 1). For birds and terrestrial mammals at the global scale, we limited our analysis to continental species because a large proportion of extinctions on oceanic islands occurred before descriptions started and sometimes even before European settlement (Harrison & Stiassny 1999;

MacPhee & Flemming 1999; IUCN 2012), preventing the use of our model. For terrestrial mammals at the global scale, we considered only the early twentieth century because our model could not fit the exponential increase in descriptions observed until that period. Focusing on that period for this group excluded 2 recorded extinctions (*Hippotragus leucopheus* and *Bos primigenius* last recorded respectively in 1800 and 1627) but only slightly changed e. We hypothesized that starting our analysis at the twentieth century for this group would, however, affect our estimate of undescribed extinctions (*NE*) because only extinctions of undescribed species in the last century would be estimated, not previous undescribed extinctions.

Exploring assumptions via simulations

We assumed that species do not differ with regard to their probability of being described or going extinct. This is, of course, a simplifying assumption that we assessed by simulating data for species that do not conform to this assumption.

Many researchers have reported that extinction or description probabilities are related to traits such as body size or geographic range size (e.g., Diniz-Filho et al. 2005). Because these biological variables are distributed approximately log normally it could be assumed that extinction and description probabilities will also conform to the log-normal distribution. However, we opted for the cloglog transformation, which is more appropriate when dealing with probabilities and very close to the log transformation when probabilities are small. To generate a pool of k species with different description probabilities, we drew random k values, $\{x_{d1}, x_{d2}, \ldots, x_{dk}\}$, from a normal distribution of mean M_d and variance V_d, and then obtained k probabilities, $\{p_{d1}, p_{d2}, \ldots, p_{dk}\}$, by using the formula $p_{di} = 1 - (1/\exp(\exp(x_{di})))$. Probabilities of extinction $\{p_{e1}, p_{e2}, \ldots, p_{ek}\}$ were obtained in the same way from a normal distribution of mean M_e and variance V_e. Non independence between p_d and p_e was set by a correlation, R, between x_d and x_e. After choosing a temporal window length, t, we used a Monte Carlo procedure to assess the description and extinction dates for each simulated species, t_d and t_e, respectively. When a species was still not described at t, t_d was arbitrarily set to $t + 1$, and if it was extant, $t_e = t + 1$. Knowing t_d and t_e allowed us to identify species that went extinct before ($t_e < t_d$, $t_d < t + 1$) or after ($t_e > t_d$, $t_e < t + 1$) being described and to construct the cumulated number of described species over the temporal window considered ($t_d < t + 1$) and hence to estimate parameters e and d from the optimization procedures described above. For each set of initial conditions (M_e, M_d, V_e, V_d, R), 30 simulations were performed. To avoid nonconvergence, high variance in the estimates, or no simulated extinctions, we considered a large number of species, namely 1000, but in a few cases (low M_e) 5000. We set the length of the temporal window at $t = 150$, a value comparable to that in the empirical data sets we analyzed.

There was a dramatic lack of information to set V_e and V_d at realistic values. To our knowledge, the variability level in description probability among species has never been quantified. Considering extinction probability, Gilpin and Diamond (1981), who have dealt with the problem of modeling interspecific probability distribution in a realistic way, concluded that a log normal distribution with unit variance was coherent with several empirical observations. Therefore, and in the absence of further information, we considered $V_e = 1$, and for the sake of simplicity we set V_d at the same value.

There are good reasons to expect a negative correlation between extinction and description probabilities. For instance, species detectability should increase with population size, whereas the opposite is expected for extinction probability. We set $R = -0.7$, a high correlation value for the interspecific relationship between 2 life-history attributes. We explored 4 situations regarding interspecific variability: variability in extinction probabilities ($V_e = 1, V_d = 0$), variability in description probabilities ($V_e = 0, V_d = 1$), and variability in both with correlation ($V_d = 1$, $V_e = 1$, $R = -0.7$) and without ($V_d = 1$, $V_e = 1$, $R = 0$). Exploratory simulations were performed to find values of M_d and M_e resulting in estimated values of d and e that covered the range observed with the empirical data sets.

Results

Unknown extinctions

The cumulative species description curves from all taxonomic groups and regions we analyzed and their fitted models (Fig. 1) showed different levels of completeness in the description record and different rates of description. Groups like birds receiving greater and earlier taxonomic efforts than others. Resulting *NE* numbers of undescribed extinct species at present (i.e., at year n, the last year of the description process, see Table 1) varied greatly among taxonomic groups, ranging from <1 to >57, respectively, for Australian birds (starting the model at year m) and North American freshwater fishes (starting the model at year $m - 25$; Table 2). These variations depended on known extinction numbers, taxonomic group size (i.e., total number of known species), starting year of extinctions, and shape of the species description curve (i.e., if species are found and described at greater rates, less time is left for unrecorded extinctions to happen). More importantly, current numbers of known extinct species were highly underestimated. On the basis of the fraction *pE* of undescribed extinct species *NE* over all extinctions (known or not), which ranged from 0.15 to 0.59

Table 2. Parameter values, numbers of undiscovered extinct species, and proportions of undescribed extinct species for each taxonomic group and region considered resulting from our model of species discovery curves.

Taxonomic group	*Region*	*Description rate* d ($sp^{-1}.yr^{-1}$)	*Projected total species richness* N_0 with $e = 0$	*Projected total species richness* N_0	*Estimated number of undescribed species extinctions NE*	*Proportion of undescribed species among all extinct species pE*
Freshwater fish	North America (U.S.A., Canada, Mexico)	0.00174 (0.00321; 0.00614*)	4263 (3046, 2264*)	4878 (3275, 2355*)	142.89 (88.18, 57.68*)	0.78 (0.68, 0.59*)
Freshwater bivalves	North America(U.S.A., Canada)	0.00206 (0.00536*; 0.00077)	392 (341*, 468)	460 (358*, 771)	20.02 (9.33*, 67.04)	0.42 (0.25*, 0.71)
Birds	world (continental)	0.02372 (0.02427*; 0.02167)	6238 (6233*, 6258)	6239 (6235*, 6260)	1.10 (1.40*, 2.05)	0.12 (0.15*, 0.20)
	Australia	0.0230439* (0.02119; 0.01706)	631* (633, 638)	632* (634, 641)	0.89* (1.14, 1.60)	0.23* (0.28, 0.35)
Amphibians	Australia	0.07256* (0.02790; 0.00442)	230* (262, 490)	233* (274, 710)	2.37* (8.44, 15.54)	0.44* (0.74, 0.84)
Marine mammals	world	0.00808 (0.00909*)	153 (150*)	158 (154*)	3.60 (3.69*)	0.44 (0.48*)
Terrestrial mammals	world (continental)	0.0174901* (0.01403; 0.00912)	4238* (4387, 4843)	4243* (4394, 4856)	4.01* (5.54, 9.29)	0.31* (0.33, 0.46)
	Australia	0.00789* (0.00752; 0.00524)	351* (365, 413)	388* (398, 469)	21.95* (25.46, 38.88)	0.54* (0.57, 0.67)

Note: Values in parentheses show the estimated values when running our model 10 and 25 years earlier than the first recorded extinction (i.e., at $t = m - 10$ and $t = m - 25$, where t is the year of the first description and m is the year of the first known extinction).

* *Values correspond to models with m values that provide the more realistic total projected richness and hence the best estimates of d (description probability), NE (number of undescribed species extinctions), and pE (proportion of undescribed species among all extinct species) (these estimates are in Fig. 1 and Results).*

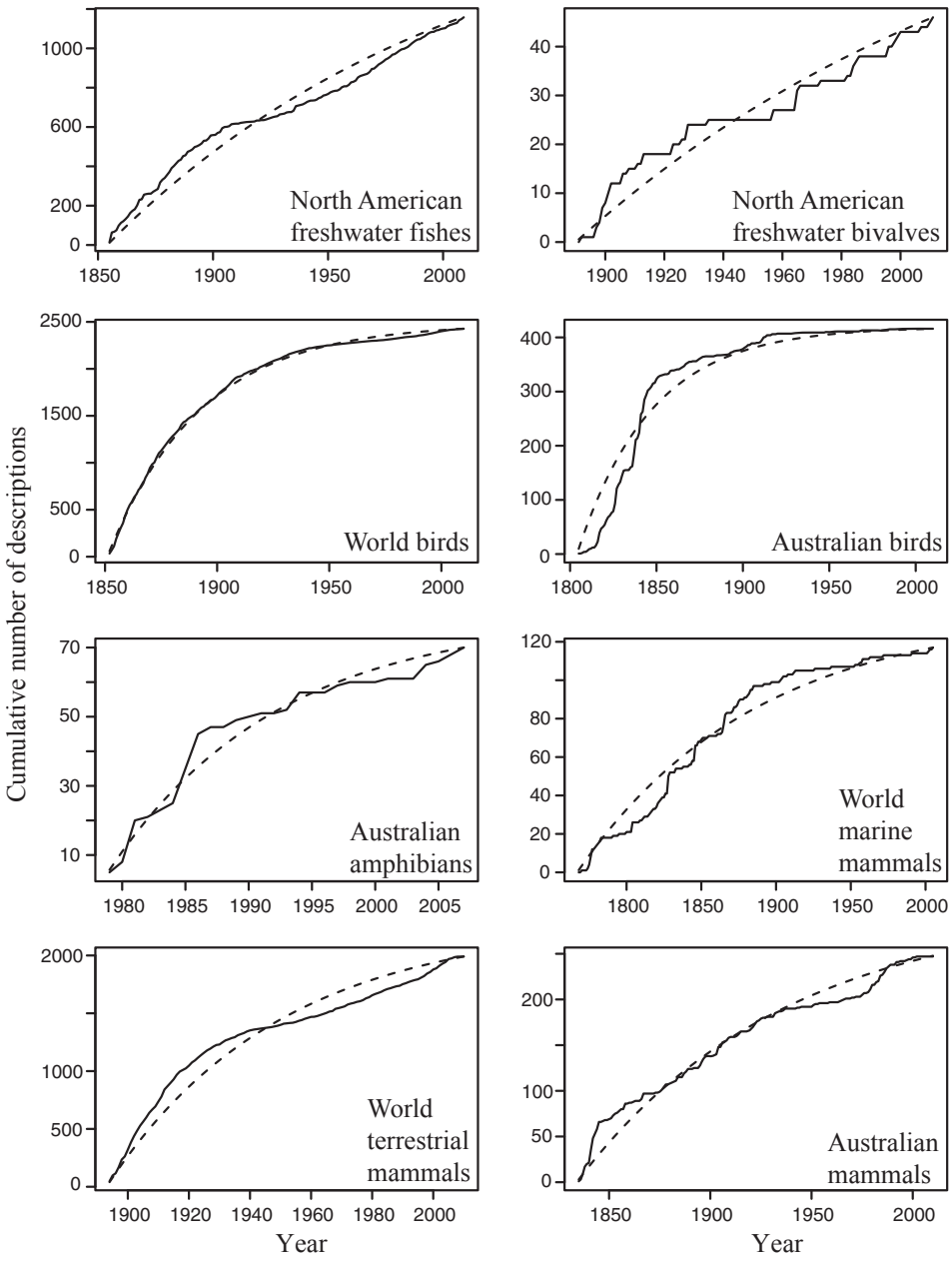

Figure 1. Cumulative number of descriptions of species for each taxonomic group and region and their fitted models (dashed lines) over time. For birds and terrestrial mammals at the global scale, we limited our analysis to continental species (i.e., excluding islands and Australia). Starting year is not always set at year m (see Table 1).

depending on the taxonomic group and region (Table 2), current extinction numbers should be nearly doubled or more in the worst cases (for North American freshwater fishes, world marine mammals, and Australian amphibians and mammals). Modifying m (i.e., the starting year of the model and of undescribed species extinctions) to 10 or 25 years earlier than the first recorded extinction altered NE values, mainly through the alteration of N_0 and d values, and in some cases yielded more realistic estimations and hence more likely values of NE (Table 2, Supporting Information). Finally, when comparing our projected total species richness with the outcomes of a model where extinctions of undescribed species were not considered (i.e., $e = 0$), projected total richness estimates (Table 2) underestimated values up to 9.7% (mean [SD] = 2.8% [3.32]).

Simulation outputs

Because both description and extinction probabilities are likely to be small, we introduced some approximations in the model. This had no effect on the estimation of N_0, e, or d, as demonstrated by the results of fitting the model to simulated data. The estimated values were close to the true values (mean absolute relative error <3%, results not shown but see Fig. 3 for pE). The approximation that resulted in Eq. 12, which neglected e in the computation

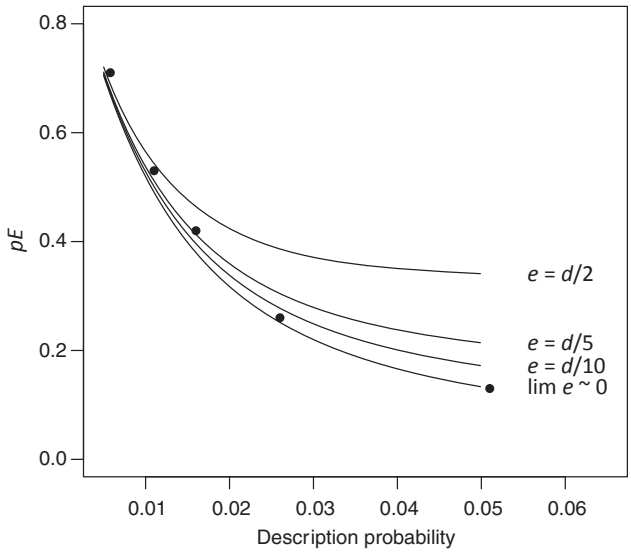

Figure 2. Predicted fraction of undescribed extinct species (pE) among all extinct species over a period of 150 years as a function of varying description probability (d) and constant extinction probability (e). The limit of the relationship when e tends toward zero is also displayed. Points are estimations of pE (see text for the estimation procedure) assuming e = 0 with simulated data (where e = 0.001, d = {0.05, 0.025, 0.015, 0.010, 0.005}, N_0 = 1000, t = 150). Each point is the average value of 30 simulations.

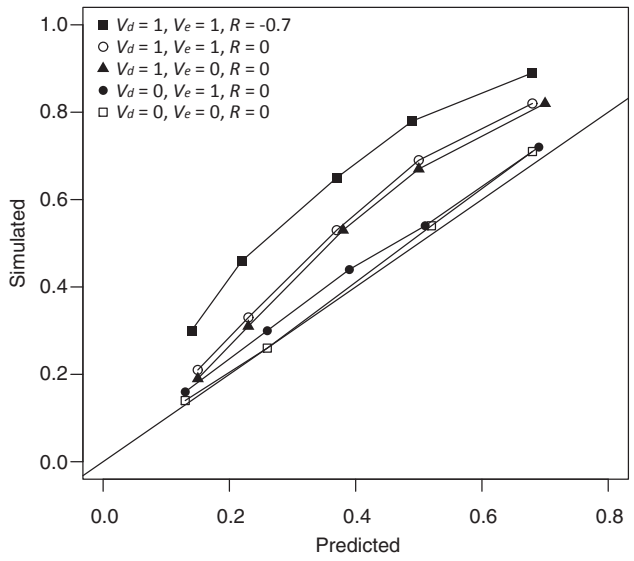

Figure 3. Simulated versus predicted values of the fraction of undescribed extinct species (pE) under different scenarios of interspecific variability in description or extinction probability (see text) (V_d and V_e, variance of description and extinction probabilities, respectively; R, correlation between both probabilities [see text]). Predicted values come from fitting a model assuming constant description and extinction probabilities (see text) to the simulated data. Each point is the average value of 30 simulations. Straight line is of perfect agreement.

of *pE*, provided acceptable results when *e* was at least an order of magnitude lower than *d* (Fig. 2).

Regarding the assumption of constant rates of description and extinction, the simulations showed that our model underestimated *pE* when species differed in their extinction or description probabilities (Fig. 3). This outcome depended mainly on the variability in description probability. When description and extinction probabilities are negatively correlated, the bias was higher than when the 2 variables were independent. As in the constant probability case, we observed that when extinction rate was on average lower than the average description rate, the fraction of undescribed species *pE* was essentially dependent on description rate.

Discussion

We applied extinction probabilities computed from recorded biodiversity losses to the proportion of undiscovered species throughout the species description process in a range of taxonomic groups. Our results showed that the current numbers of species officially considered extinct in nature were highly underestimated values of the actual number of extinctions. The proportion of undiscovered extinct species over all extinctions ranged from 0.15 to 0.59, depending on the taxonomic group and the region considered. This means that today's known extinctions should be more than doubled in some cases. Although unnoticed species extinctions have been suggested as a logical consequence of ongoing human alteration of natural systems (Ceballos & Ehrlich 2009; Giam et al. 2012), ours was the first attempt to quantify unknown diversity loss. (We considered equal extinction probabilities for both described an undescribed species and quantified net diversity loss, not change in extinction rate.)

We assumed that the probability of describing a species is constant through time, an unrealistic assumption in the presence of interspecific variability. Species with high description probability are likely to be described first, whereas species with low probability will be described later; thus, the average description probability of species still undescribed decreases over time. A similar reasoning holds for extinction. Therefore, such time dependent rates are likely to have biased our results. Our simulations accounting for interspecific variability support this view. Our model underestimated undescribed extinct species numbers by about 10% or even 20% when description and extinction probabilities were negatively correlated (Fig. 3). This negative correlation translates to the likely

assumption that extinction rates of undiscovered species should increase over time because the species remaining to be discovered are more likely to be range restricted. Indeed, range size is the best single predictor of extinction risk (Purvis et al. 2000*a*), and it is used as a metric to designate current species threats by the IUCN (2012). Giam et al. (2012) observed a positive relationship between extinction proneness of species and year of description, suggesting that undescribed species are in increasing danger of extinction as time goes on. The negative relationship between extinction probability and detectability is likely to bias our results toward low values of undescribed extinct species and should hence be considered as a first conservative estimate. Interspecific variability also leads to biased estimates of the number of undescribed species and is a potential source of concern for all researchers who aim to quantify unknown biodiversity. Knowledge about the level of interspecific variability in extinction and description rates is at present so limited that it is difficult to assess whether or not these differences could be neglected.

The fraction of extinct species that have not been described is mainly driven by description rate, and future efforts should be devoted to finding a better way to estimate this parameter. For instance, our model could be made more complex by integrating effort as it relates to taxonomic group (e.g., variation in the number of descriptions by author) in the description process (Joppa et al. 2011*a*). Inclusion of effort could lead to a better model fit and hence more accurate parameter estimations. However, because we focused on relatively well-known groups and our model fitted the description curves relatively well (Fig. 1), applying a more complex model would likely only slightly modify our results. Similarly, the likelihood of concluding that a species is globally extinct depends on the effort devoted to finding it, which may vary in space and time. It is likely that some species described a long time ago and known only from their type specimens are extinct by now. In contrast, some species formerly considered extinct may be delisted following intensive field work. Integrating the history of sampling effort (Solow 2005), such as numbers of specimens housed in scientific institutions, into our approach could lead to improved estimates of extinction probabilities.

Our conclusion that the number of recorded extinctions substantially underestimated true extinctions is essentially bad news, particularly for understudied taxonomic groups or regions because the bias decreased with description rate. But our results also bring good news. First, for every case we considered, the comparison of extinction and description probabilities led to the reassuring conclusion that an undescribed species is far more likely to be described than to go extinct in the near future (Costello et al. 2013). However, our method is based on the assumption that description curves are convex, which might not be representative for less well-known taxonomic groups or regions. Second, our results showed that assuming $e = 0$ (i.e., neglecting the extinction of undescribed species in modeling the discovery record), which is a common assumption for all models attempting to predict total species richness (e.g., Solow & Smith 2005; Bebber et al. 2007; Joppa et al. 2011*a*; Mora et al. 2011), only slightly underestimated this total projected richness. Although this underestimation can reach over 9% for one of the most affected taxonomic groups (Australian mammals), these findings support previous predictions of total species richness. Under some realistic assumptions, even the fraction of undescribed extinct species could be estimated without having to estimate e, an interesting option when the number of observed extinctions is low or not known with great confidence, a common situation.

In the light of our results, what should be the strategy of conservation professionals? It goes without saying that they should try to preserve as many species as possible. However, it is not intellectually simple to justify and ensure the preservation of yet undiscovered diversity. As stated by Ceballos and Ehrlich (2009), "the more diversity that is discovered the more urgent becomes putting additional resources into understanding and finding ways to conserving [*sic*] it." We reached a somewhat opposite conclusion: the more diversity still to be discovered, the more urgent it becomes to put additional resources into understanding and finding ways to conserve it. Knowing that approximately 90% of existing species await description (Mora et al. 2011), our results are an urgent call for ecosystem conservation and taxonomic research. Conserving natural systems where large proportions of biodiversity are yet to be discovered, without focusing on particular species, seems the most reasonable way to give unknown species a chance to be discovered before they become extinct and to allow for planning of their conservation. Indeed, to obtain financial support to protect a species from human threats, the species first has to be discovered alive. The increasing number of taxonomists working on large-sized taxa (Joppa et al. 2011*b*) may reflect their increase in developing countries (Costello et al. 2013), where most biodiversity hotspots are located. Although this gives us good reasons to be optimistic for the discovery and conservation of the taxonomic groups analyzed here, we should keep in mind that the majority of organisms are invertebrates, fungi, and microorganisms, which are still largely undescribed and likely to remain so if the number of taxonomists specializing in these organisms does not increase.

Acknowledgments

We thank the French National Agency for Research (ANR-09-PEXT-008) and the European Commission (BioFresh

project: FP7-ENV-2008-Contract n°226874) for financial support. X.G. was supported by the Princeton University Graduate Fellowship and the Wildlife Reserves Singapore Conservation Fund. We are very grateful to BirdLife International and IUCN for providing species lists and geographic data. P.A.T., A.E.B, and R.B. also thank the scientists' network specializing in freshwater biodiversity (Freshwater Animal Diversity Assessment) for providing a framework and space to meet and develop new data and ideas. We also thank 2 anonymous referees and the associate editor for helpful comments that greatly improved the manuscript.

Supporting Information

A detailed explanation of how the approximation of parameter *e* was estimated (Appendix S1), information on species considered extinct, bibliographic source, former distribution, and estimated extinction date (Appendix S2), and the description trends and fitted models applying the 3 different possible starting years (m, $m - 10$, and $m - 25$) (Appendix S3) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

Barnosky, A. D., et al. 2011. Has the Earth's sixth mass extinction already arrived? Nature **471:**51–57.

Bebber, D. P., F. H. Marriott, K. J. Gaston, S. A. Harris, and R. W. Scotland. 2007. Predicting unknown species numbers using discovery curves. Proceedings of the Royal Society B: Biological Sciences **274:**1651–1658.

BirdLife International, and NatureServe. 2011. Bird species distribution maps of the world.

Bogan, A. E. 1993. Freshwater Bivalve Extinctions (Mollusca: Unionoida): a search for causes. American Zoologist **33:**599–609.

Bogan, A. E. 1998. Freshwater molluscan conservation in North America: problems and practices. Pages 223–230 in I. J. Killeen, M. B. Seddon, A. M. Holmes, editors. Molluscan conservation: a strategy for the 21st Century. Journal of Conchology, Special Publication.

Burbidge, A. A., N. L. McKenzie, K. E. C. Brennan, J. C. Z. Woinarski, C. R. Dickman, A. Baynes, G. Gordon, P. W. Menkhorst, and A. C. Robinson. 2009. Conservation status and biogeography of Australia's terrestrial mammals. Australian Journal of Zoology **56:**411–422.

Ceballos, G., and P. R. Ehrlich. 2009. Discoveries of new mammal species and their implications for conservation and ecosystem services. Proceedings of the National Academy of Sciences **106:**3841–3847.

Costello, M. J., R. M. May, and N. E. Stork. 2013. Can we name Earth's species before they go extinct? Science **339:**413–416.

Diniz-Filho, J. A. F., R. P. Bastos, T. F. L. V. B. Rangel, L. M. Bini, P. Carvalho, and R. J. Silva. 2005. Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. Global Ecology and Biogeography **14:**469–477.

Giam, X., T. H. Ng, V. B. Yap, and H. T. W. Tan. 2010. The extent of undiscovered species in Southeast Asia. Biodiversity and Conservation **19:**943–954.

Giam, X., B. R. Scheffers, N. S. Sodhi, D. S. Wilcove, G. Ceballos, and P. R. Ehrlich. 2012. Reservoirs of richness: least disturbed tropical forests are centres of undescribed species diversity. Proceedings of the Royal Society B: Biological Sciences **279:**67–76.

Gilpin, M. E., and J. M. Diamond. 1981. Immigration and extinction probabilities for individual species: relation to incidence functions and species colonization curves. Proceedings of the National Academy of Sciences of the United States of America **78:**392–396.

Haag, W. R. 2009. Past and future patterns of freshwater mussel extinctions in North America during the Holocene. Pages 107–128 in S. Turvey, editor. Holocene extinctions. Oxford University Press, United Kingdom.

Harrison, I. J., and M. L. J. Stiassny. 1999. The quiet crisis: a preliminary listing of the freshwater fishes of the world that are extinct or "missing in action" (database: http://www.creo.amnh.org). Pages 271–332 in R. D. E. MacPhee, editor. Extinctions in near time: causes, contexts, and consequences. Kluwer Academic/Plenum Publishers, New York.

IUCN. 2012. The IUCN red list of threatened species. Version 2012.1. Available from: <http://www.iucnredlist.org>.

Jelks, H. L., et al. 2008. Conservation status of imperiled North American freshwater and diadromous fishes. Fisheries **33:**372–407.

Joppa, L. N., D. L. Roberts, and S. L. Pimm. 2011*a*. How many species of flowering plants are there? Proceedings of the Royal Society B: Biological Sciences **278:**554–559.

Joppa, L. N., D. L. Roberts, and S. L. Pimm. 2011*b*. The population ecology and social behaviour of taxonomists. Trends in Ecology & Evolution **26:**551–553.

MacPhee, R. D. E., and C. Flemming. 1999. Requiem Æternam: the last five hundred years of mammalian species extinctions (database: http://www.creo.amnh.org). Pages 333–372 in R. D. E. MacPhee, editor. Extinctions in near time: causes, contexts, and consequences. Kluwer Academic/Plenum Publishers, New York.

McKinney, M. L. 1997. Extinction vulnerability and selectivity: combining ecological and paleontological views. Annual Reviews in Ecology and Systematics **28:**495–516.

Miller, R. R. 2005. Freshwater fishes of Mexico. The University of Chicago Press, Chicago.

Mora, C., D. P. Tittensor, S. Adl, A. G. Simpson, and B. Worm. 2011. How many species are there on Earth and in the ocean? PLoS Biology **9:**e1001127.

Pereira, H. M., et al. 2010. Scenarios for global biodiversity in the 21st century. Science **330:**1496–1501.

Pimm, S. L., M. P. Moulton, L. J. Justice, N. J. Collar, D. M. J. S. Bowman, and W. J. Bond. 1994. Bird Extinctions in the Central Pacific [and Discussion]. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **344:**27–33.

Purvis, A., J. L. Gittleman, G. Cowlishaw, and G. M. Mace. 2000*a*. Predicting extinction risk in declining species. Proceedings of the Royal Society B: Biological Sciences **267:**1947–1952.

Purvis, A., K. E. Jones, and G. M. Mace. 2000*b*. Extinction. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology **22:**1123–1133.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/.

Régnier, C., B. Fontaine, and P. Bouchet. 2009. Not knowing, not recording, not listing: numerous unnoticed mollusk extinctions. Conservation Biology **23:**1214–1221.

Richling, I., and P. Bouchet. 2013. Extinct even before scientific recognition: a remarkable radiation of helicinid snails (Helicinidae) on the

Gambier Islands, French Polynesia. Biodiversity and Conservation **22:**2433–2468.

Scheffers, B. R., L. N. Joppa, S. L. Pimm, and W. F. Laurance. 2012. What we know and don't know about Earth's missing biodiversity. Trends in Ecology & Evolution **27:**501–510.

Solow, A. R. 2005. Inferring extinction from a sighting record. Mathematical Biosciences **195:**47–55.

Solow, A. R., and W. K. Smith. 2005. On estimating the number of species from the discovery record. Proceedings of the Royal Society B: Biological Sciences **272:**285–287.

Stork, N. E. 1993. How many species are there? Biodiversity and Conservation **2:**215–232.

Stork, N. E. 2010. Re-assessing current extinction rates. Biodiversity and Conservation **19:**357–371.

Tedesco, P. A., C. Lévêque, T. Oberdorff, D. Paugy, and M. L. J. Stiassny. 2010. World checklist of freshwater fish species. Available from: http://fada.biodiversity.be/group/show/52.

Turgeon, D. D., et al. 1998. Common and scientific names of aquatic invertebrates from the United States and Canada. 2nd ed. American Fisheries Society, Bethesda, Maryland.